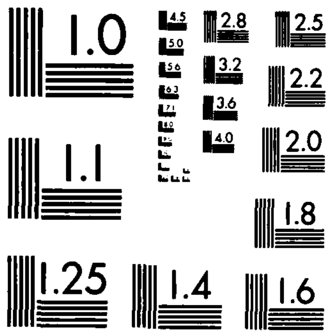
AD-A165 290 STATISTICAL SIGNIFICANCE AND BASELINE MONITORING(U) OLD DOMINION UNIV NORFOLK VA APPLIED MARINE RESEARCH LAB R W ALDEN JUL 84 DACW65-81-C-0051 1/1

UNCLASSIFIED F/G 6/6 NL


1.0
4.5
5.0
5.6
6.3
7.1
8.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

OLD DOMINION UNIVERSITY RESEARCH FOUNDATION

AD-A165 290

APPLIED MARINE RESEARCH LABORATORY
OLD DOMINION UNIVERSITY
NORFOLK, VIRGINIA

STATISTICAL SIGNIFICANCE AND BASELINE MONITORING

By

Raymond W. Alden, Principal Investigator

Supplemental Contract Report
For the period ending September 1984

Prepared for the
Department of the Army
Norfolk District, Corps of Engineers
Fort Norfolk, 803 Front Street
Norfolk, Virginia 23510

Under
Contract DACW65-81-C-0051
Work Order No. 0016

US Army Corps
Of Engineers
Norfolk District

July 1984

Report B- 20

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

AD-A165290

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release, distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
B-20

6a. NAME OF PERFORMING ORGANIZATION
Old Dominion University, Applied Marine Research Lab.

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
U.S. Army Corps of Engineers, Norfolk District

6c. ADDRESS *(City, State, and ZIP Code)*
Norfolk, Virginia 23508

7b. ADDRESS *(City, State, and ZIP Code)*
Norfolk, Virginia 23510-1096

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
U.S. Army Corps of Engineers, Norfolk District

8b. OFFICE SYMBOL *(If applicable)*
NAOPL; NAOEN

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DACW65-8[illegible]-C-0051

8c. ADDRESS *(City, State, and ZIP Code)*
Norfolk, Virginia 23510-1096

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
Statistical Significance and Baseline Monitoring

12. PERSONAL AUTHOR(S)
Alden, R.W., III

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM ______ TO ______	1984, July	29

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
statistical analysis, impact significance, baseline monitoring, trend assessment, ecological impact, monitoring program, minimum detectable impacts (MDIs)

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*
The effectiveness of several multivariate techniques for the detection of impacts in data from trend assessment studies was evaluated. PCA models proved to be not very sensitive; discriminant analysis were overly sensitive; MANOVA techniques proved to be the most effective in detecting significant change.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Craig L. Seltzer

22b. TELEPHONE *(Include Area Code)*
(804) 441-3767/827-3767

22c. OFFICE SYMBOL
NAOPL-R

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

APPLIED MARINE RESEARCH LABORATORY
OLD DOMINION UNIVERSITY
NORFOLK, VIRGINIA

STATISTICAL SIGNIFICANCE AND BASELINE MONITORING

By

Raymond W. Alden, Principal Investigator

Supplemental Contract Report
For the period ending September 1984

Prepared for the
Department of the Army
Norfolk District, Corps of Engineers
Fort Norfolk, 803 Front Street
Norfolk, Virginia 23510

Under
Contract DACW65-81-C-0051
Work Order No. 0016

Submitted by the
Old Dominion University Research Foundation
P. O. Box 6369
Norfolk, Virginia 23508

July 1984

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES - Concluded

STATISTICAL SIGNIFICANCE AND BASELINE MONITORING

By

Raymond W. Alden*

INTRODUCTION

Scientists developing environmental monitoring programs must consider the ultimate question: Has a significant impact occurred? This question represents a primary concern of environmentalists and regulatory agencies alike. Therefore, the investigator must design baseline and trend assessment studies in such a way as to allow detection of environmental impacts. However, to properly address this question the investigator should have a basic definition of a "significant" impact. Typically, it is approached as two components: 1) What is a statistical impact? and 2) What is an ecological impact? In order for an impact to be considered ecologically significant, it really should be statistically significant. However, the converse is not necessarily true. In fact, it would be desirable to design a monitoring program which would allow the statistical detection of ecological changes before they become critical.

~~The present study~~ offers strategies for defining statistical impacts for an environmental monitoring program. Specifically, a series of statistical techniques have been developed to estimate "minimum detectable impacts" (MDIs) for variables examined during the baseline phase of a monitoring program at an open ocean dredged material disposal site. The MDIs are dependent upon natural spatiotemporal variability of baseline data and the intensity of the monitoring effort.

*Director, Applied Marine Research Laboratory, Old Dominion University, Norfolk, Virginia 23508.

EVALUATION AND DISCUSSION OF METHODS

MDIs of Single Samples

One of the most basic topics which may be considered concerns how different a single sample must be before a statistically significant impact can be inferred. Green (1979) considers a number of statistical methods for evaluating environmental data. One approach which Green suggests to detect statistically significant outliers involves the evaluation of the samples against the context of the variance-covariance relationships of the baseline data set. The method involves the use of a chi-square (χ^2) test of a sample of variables employing the following equation:

$$\chi^2_{(pdf)} = (\hat{X}_j - \bar{X}_j) D^{-1} (\hat{X}_j - \bar{X}_j) \qquad (1)$$

where $\hat{X}_j$ is the value of the new observation; $\bar{X}_j$ is the mean or expected value of the j^{th} variable, and D is the variance-covarience matrix.

If the sample being tested is sufficiently divergent from the original data set, equation (1) will produce a χ^2 value greater than the critical test level for p degrees of freedom. This equation can also be used to predict the MDI levels for single samples.

A computer program was developed to add or subtract factors (i.e. small percentages of the means) incrementally to the means of each variable. This program was coupled with the χ^2 test (equation 1) to evaluate iteratively the effects of increasing or decreasing the values of each variable. Each variable was evaluated sequentially, with all other variables remaining constant and equal to the means. The factors were changed incrementally for the variable being tested until a significant χ^2 value was attained. The values of the factors which just produced significant χ^2 values for each variable were considered to be the MDI levels.

To provide an example of how this method may be used, a subset of data from a baseline water quality monitoring program at a potential open ocean dredged material disposal site was subjected to the MDI evaluation process. Data from two cruises taken in late summer (August) and early autumn (October) were selected from a three year baseline program (1981-1983). These data would be expected to exhibit the sort of natural spatiotemporal variability which may be observed for any given seasonal period. The means, the values of the simulated "impacted" sample, and the MDI factors estimated for each of the water quality variables are presented in Table I. Parameters which may be expected to decrease when impacted by dredged material disposal operations (e.g. dissolved oxygen, pH, plant pigments) had the factors subtracted from the means, while those which would be expected to increase had the factors added.

Although there was a fairly wide range of MDI levels, from 5% to over 400%, none of the "impacted" values were at a level considered to be extremely harmful ecologically. The parameters with the greatest relative MDIs were the ones which were found at extremely low levels, often near detection limits. Therefore, the absolute concentrations of the "impacted" samples, although significantly different from the baseline means, were still moderately low. In fact, few of the "impacted" values fell outside of the natural range reported by Kester and Courant (1973) for estuarine Chesapeake Bay waters, and none approached the water quality criteria or reference levels recommended by state and federal agencies for the protection of marine life, or the prevention of eutrophication (Virginia State Water Control Board, 1976). Thus, the approach allowed the detection of a

Table I. Results of MDI analysis for single samples. Means, standard errors, estimated values of the "impacted" sample and MDI levels are presented for each parameter.

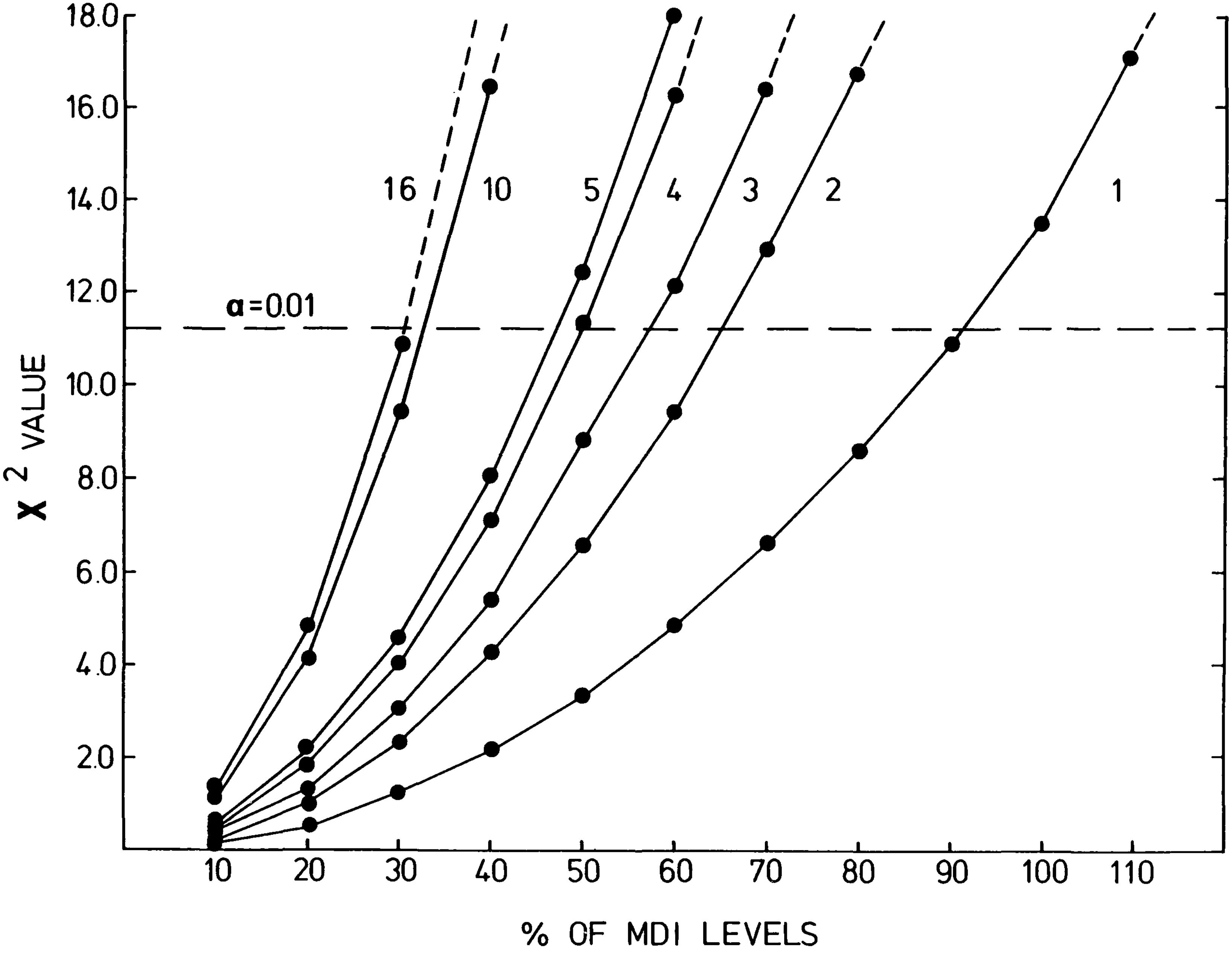
VARIABLE	UNIT	MEAN	STANDARD ERRORS	"IMPACTED" VALUE	M.D.I. (% OF MEAN)
Dissolved Oxygen	mg/l	8.78	0.21	6.15	- 30
pH	--	7.85	0.01	7.46	- 5
C.O.D.	mg/l	30.60	1.49	7.46	+105
Turbidity	NTU	1.40	0.10	2.03	+145
Nitrate-Nitrite	μg/l	0.30	0.07	1.53	+410
Orthophosphate	mg/l	0.002	0.0005	0.009	+352
Total Phosphorus	mg/l	0.009	0.002	0.035	+290
TKN	mg/l	0.133	0.005	0.259	+ 95
Ammonia	mg/l	0.092	0.005	0.212	+130
Suspended Solids	mg/l	12.01	0.59	22.22	+ 85
Volatile Nonfilterable Residue	mg/l	3.29	0.21	6.91	+110
Chlorophyll _a_	mg/l	4.31	0.14	3.88	- 10
Chlorophyll _a_ (corrected)	mg/l	4.25	0.15	3.61	- 15
Chlorophyll _b_	mg/l	0.51	0.02	0.18	- 65
Chlorophyll _c_	mg/l	2.19	0.09	1.31	- 40
Phaeophytin	mg/l	0.28	0.06	0.88	+215

statistically significant impact at a level below that which may be of acute ecological significance. Of course, this is the desired situation if the trend assessment is to act as an "early warning system" for the detection of an impact before the environment deteriorates excessively.

It is anticipated that during the major environmental perturbation more than one variable is impacted. The amount of change in a given variable required for statistical detection would then be expected to be considerably less. An empirical examination of the same data base was tested with various impact levels being introduced. Each testing series consisted of the evaluation of data sets for which the values of various combinations of variables were changed by factors from 10% to 100% of the MDI levels. Series were tested for 1, 2, 3, 4, 5, 10 and 16 variables being changed at the same time. For the series involving 1, 2, and 16 variables, all possible combinations were tested at each impact level. Since the numbers of permutations of variables in the other series were too large to allow the examination of all combinations, ten randomly selected combinations were tested.

The mean χ^2 values for each impact level tested for each series are presented in Figure 1. The most obvious trend is that, the more variables being changed, the smaller the impact required for any given variable to produce a statistically significant χ^2 value. In fact, for virtually all cases involving impacts to multiple variables, changes much greater than 60% of the MDI produce signficant χ^2 values. On the other hand, tests involving impacts of 30% of the MDI or less do not produce significant χ^2 values, even when all variables are involved. Therefore, the MDI factors are conservative estimates of statistically significant impacts which appear to be within a range of two to three times the levels causing an impact when more than one variable is affected.

Figure 1. Mean χ^2 value versus the percent of the MDI's for a series of emperical tests. The data points for series involving 1, 2 and 16 variables being impacted at once represent the means of all possible combinations, while those for the other series (3, 4, 5 and 10 variables impacted at once) represent the means of tests of ten randomly selected combinations for each impact level.


18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
χ^2 VALUE
α=0.01
16
10
5
4
3
2
1
10
20
30
40
50
60
70
80
90
100
110
% OF MDI LEVELS

MDIs For Data Sets

The most common circumstances for the examination of data for impacts involves the collecti n of multiple samples in a trend assessment study. Therefore, it is desirable to develop statistical models for the evaluation of entire post-impact data set (e.g. water quality data from a cruise taken after a disposal site becomes active).

Multivariate statistical models are very useful in the determination of patterns in baseline data, as well as in the detection of impacts in trend assessment data (Alden et al., 1982). Such tests provide a single answer concerning the data set and avoid the multiple tests required by the univariate approach. Multiple univariate tests in the context of a monitoring program mean that numerous false alarms (i.e. Type I errors) crop up over time by chance alone. Therefore, the models for post-impact data sets are based upon various multivariate statistical approaches.

Statistical models for trend assessment studies may fall into several major categories: baseline, seasonal, and seasonal-area interaction contrasts. The baseline contrast models evaluate differences between the data from a post-impact collection and the entire baseline data set. The seasonal models compare the post-impact data with data collected during the same season under baseline conditions. The seasonal models are generally more sensitive than the baseline models because natural season to season variations are not included in the error terms of the seasonal contrasts.

The approach which is potentially the most sensitive at detecting impacts, involves the seasonal-area interaction models. This type of model assumes that there is an area which, during the post-impact period, exhibits conditions which would be identical to those in the impacted area, if a perturbation had not occurred. Therefore, seasonal and aperiodic temporal variations are "filtered out" statistically before impacts are evaluated. Green (1979) describes various aspects of these models in detail.

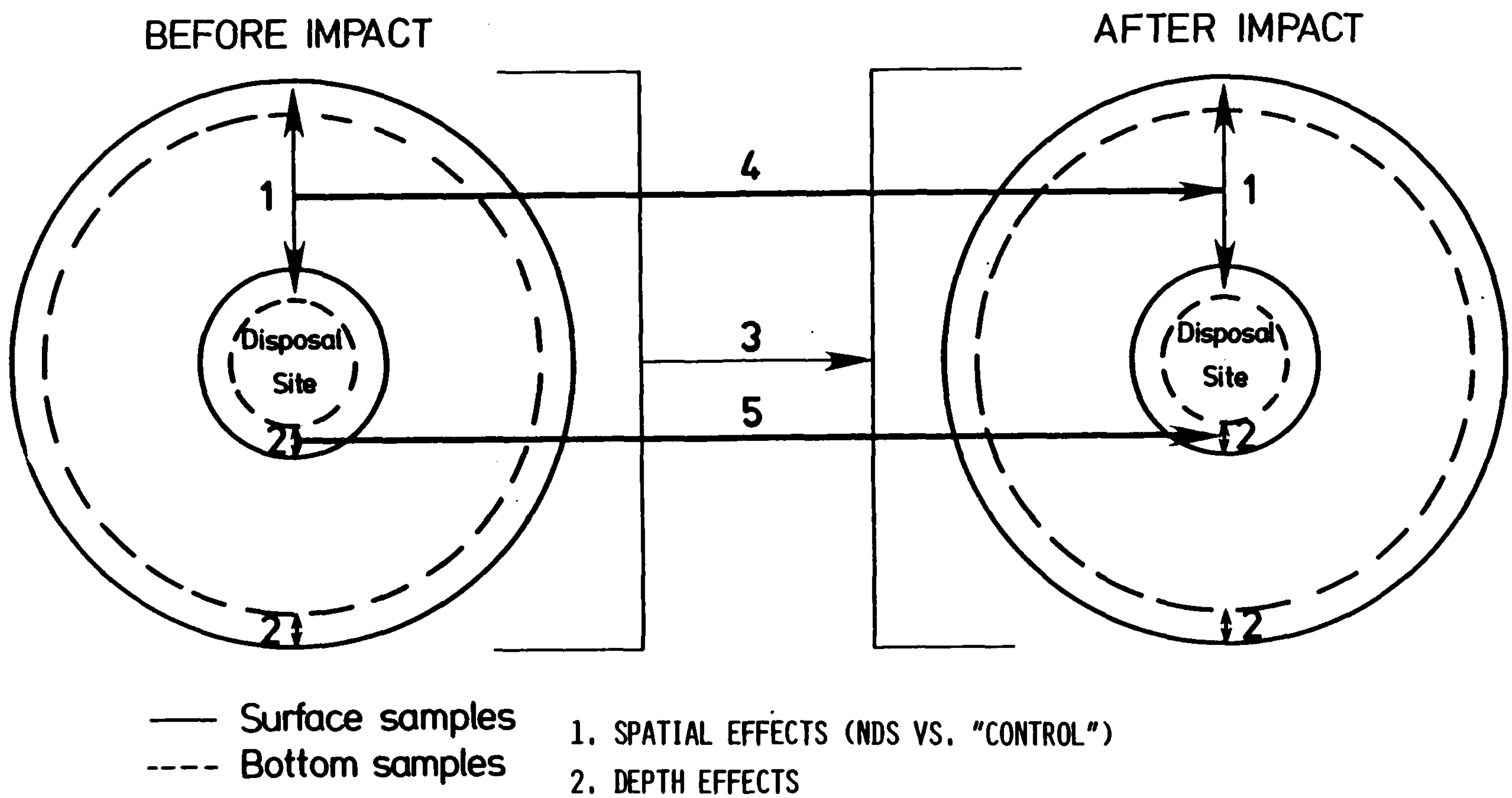
The types of effects schematically tested in a seasonal-area interaction best approach the study of an open ocean disposal site (The Norfolk Disposal Site - "NDS") (Figure 2). Effect 1 represents the comparison between water quality at the disposal site and that of surrounding waters. Effect 2 is the difference between surface and bottom water samples. The third effect is the expected seasonal change. The fourth and fifth effects are the responses of interest: those impacts caused by disposal operations at the site. By correcting for Effects 1-3 prior to the statistical evaluation of impact effects, much of the natural spatiotemporal variations can be eliminated from the analysis.

As with the case of the single sample, the MDI's for data sets are expressed as the factors which just produce statistically significant results in post-impact data sets. The MDIs are proportions of the means of the variables which may be positive (for enhancement) or negative (for inhibition), depending upon the expected responses to the disposal operations. Factors are added iteratively to the variables in the unimpacted data set until a statistically significant test criterion results for each.

Two groups are compared with the baseline and seasonal models: the reference or control groups, and the "impacted" data set. Therefore, the MDIs for the variables are based upon the F-value from a single classification ANOVA model for unequal sample sizes. The F-value test criterion is defined as:

$$F_{[1,n-a]} = (SS_G \div (a-1)) \div MS_W \qquad (2)$$

Figure 2. Schematic presentation of the effects tested in seasonal-area interaction models. The inner circles represent data from the disposal site, while the other circles represent the "control" data from surrounding areas. The solid line circles represent surface water data, while the broken line circles represent bottom water data. The arrows represent effects being tested. (See text for more detailed explanation).



1. SPATIAL EFFECTS (NDS VS. "CONTROL")
2. DEPTH EFFECTS
3. TIME EFFECTS
4. TIME-DISPOSAL SITE INTERACTION
   (I.E., CHANGES AT NDS NOT OBSERVED ELSEWHERE)
5. TIME-DISPOSAL SITE-DEPTH INTERACTION
   (I.E., CHANGES AT NDS BOTTOM SAMPLES NOT OBSERVED ELSEWHERE)

where n is the total number of samples, $a = 2$ for the two groups, and MS_W is the variance of the groups. SS_G is the sum of the squares for groups, and is defined as:

$$SS_G = (((\overline{X}_B)^2 x R_B)+((\overline{X}_A)^2 x R_A))-(((\overline{X}_B x R_B)+(\overline{X}_A x R_A))^2 \ (R_B+R_A)) \qquad (3)$$

where $\overline{X}_B$ and $\overline{X}_A$ are the mean values for variables in the before and after data, respectively, and R_B and R_A are the number of replicates in the two sets. The $\overline{X}_B$ is calculated for the seasonal or baseline data set, while the $\overline{X}_A$ is the mean for a data set to which the incremented factors have been added iteratively until a significant F-value is produced.

The method for calculating the MDIs for a season-area interaction model is somewhat more complicated, but is based upon a similar approach. Green (1979) presents a method for the determination of the number of replicates required in a sampling regime to statistically detect a given level of change in the variables between a "control" and an "impacted" area. The method for determining MDIs essentially examines the converse of this situation: the level of a statistically significant "impact" is determined for a given sampling regime (known number of replicates) and a predetermined level of spatiotemporal variability. Using the terms employed by Green (1979) and omitting several steps, the test criterion for the interaction effect is:

$$F_{[1,\varepsilon(R-1)]} = SS_{INT} \div MS_W = ((((\overline{Z}_{AC} x R_{AC})+(\overline{Z}_{BI} x R_{BI}))-((\overline{Z}_{BC} x R_{BC})(\overline{Z}_{AI} x R_{AI})))^2$$

$$\div (R_{AC}+R_{BI}+R_{BC}+R_{AI})) \div S_Z^2 \qquad (4)$$

where SS_{INT} is the interaction sum of squares; MS_W is the error or within group mean square which is assumed to be equal to S_Z^2, the variance of the groups; $\bar{Z}_{BC}$, $\bar{Z}_{BI}$, $\bar{Z}_{AC}$, and $\bar{Z}_{AI}$ are the means for the control and impact areas before and after perturbation; and R_{BC}, R_{BI}, R_{AC} and R_{AI} are the replicates for the various groups. Green assumed that $\bar{Z}_{BC} = \bar{Z}_{BI} = \bar{Z}_{AC}$, and that $\bar{Z}_{AI} = \bar{Z}_{BI}$ - (impact factor); the variance term for all groups equaled S_Z^2; and the number of replicates R, the term for which he solved for, was equal in all groups.

The mean value in the MDI method for each group as well as a pooled variance term is calculated. The actual number of replicates for each group is also used. Since MDIs are calculated prior to an actual impact, $\bar{Z}_{AC} \approx \bar{Z}_{AI}$, so an "impact" is introduced: $\bar{Z}_{AI*} = \bar{Z}_{AI} + (\bar{Z}_{AI} \times \text{impact factor})$. The $\bar{Z}_{AI*}$ values are adjusted by incremental changes in the impact factor until a significant F-value is produced through interactive tests employing equation 4. When a significant interaction term is produced, the impact factor becomes the MDI for the variable being tested. The computer program for evaluating MDI's accepts data sets from the disposal site and surrounding water from two seasons and sequentially evaluates the variables to produce a MDI value for each. Thus, a data set with a great deal of natural spatiotemporal variations and relatively low level of replication will necessitate a greater level of change before statistical significance is achieved (i.e. a set of larger MDI values is produced).

The MDI values for the baseline, seasonal and seasonal-area interaction models for data from two (August and October) water quality cruises taken as part of a baseline monitoring cruise to the Norfolk Disposal Site are presented in Table II. In general, the MDI values for the variables under the baseline model are greater than those for the seasonal model, which, in

Table II. Results of MDI analysis for baseline, seasonal and seasonal-area interaction models.

PREDICTED MDI'S (% OF MEANS)

PARAMETER	BASELINE	SEASONAL	SEASON-AREA INTERACTION
Dissolved oxygen	-30%	-30%	-15%
pH	-5%	-5%	-5%
Chemical Oxygen Demand	25%	25%	25%
Turbidity	120%	120%	120%
Nitrate-Nitrite	325%	210%	*
Orthophosphate	800%	400%	200%
Total Phosphorous	<3000%	<3000%	200%
Total Kjeldahl Nitrogen	230%	10%	5%
Ammonia	60%	45%	35%
Suspended Solids	15%	10%	15%
Volatile Nonfilterable Residue	15%	15%	10%
Chlorophyll a	-25%	-20%	-25%
Chlorophyll a (corrected)	-25%	-25%	-25%
Chlorophyll b	-45%	-15%	-15%
Chlorophyll c	-40%	-5%	-5%
Phaeophytin	700%	300%	111%

*At least one group of values in the model were all below detection limits.

turn, are larger than those for the seasonal-area interaction model. The pattern is especially apparent for the parameters which display the greatest variability, and thus have the greatest MDI levels. This trend is obviously due to the fact that more of the spatiotemporal variability is accounted for by the interaction model. However, the greater degrees of freedom in the baseline model and, to a lesser extent, in the seasonal model partially offset this pattern. The trend would be greater if year-to-year variations are included (i.e. the evaluation of two years of baseline data, or the data from the same seasonal period on different years). The seasonal-area model accounts for this source of variation as long as a natural change does not occur in one of the areas that is not also taking place in the other. Presumably, the amount of added variation observed in a continuing baseline monitoring program would become somewhat asymptotic over time, so the MDI levels should stabilize. This trend would be useful to analyze in order to determine the level of effort required to "get a handle" on natural spatiotemporal variability prior to trend assessment studies. Graphically or statistically relating the MDI values to the cumulative sampling effort (e.g. the period, number of cruises or the number of samples in the data set) may provide useful insight into the design of effective baseline programs. Of course, long-term natural trends (e.g. regional degradation in water quality over decades) would tend to confound these findings.

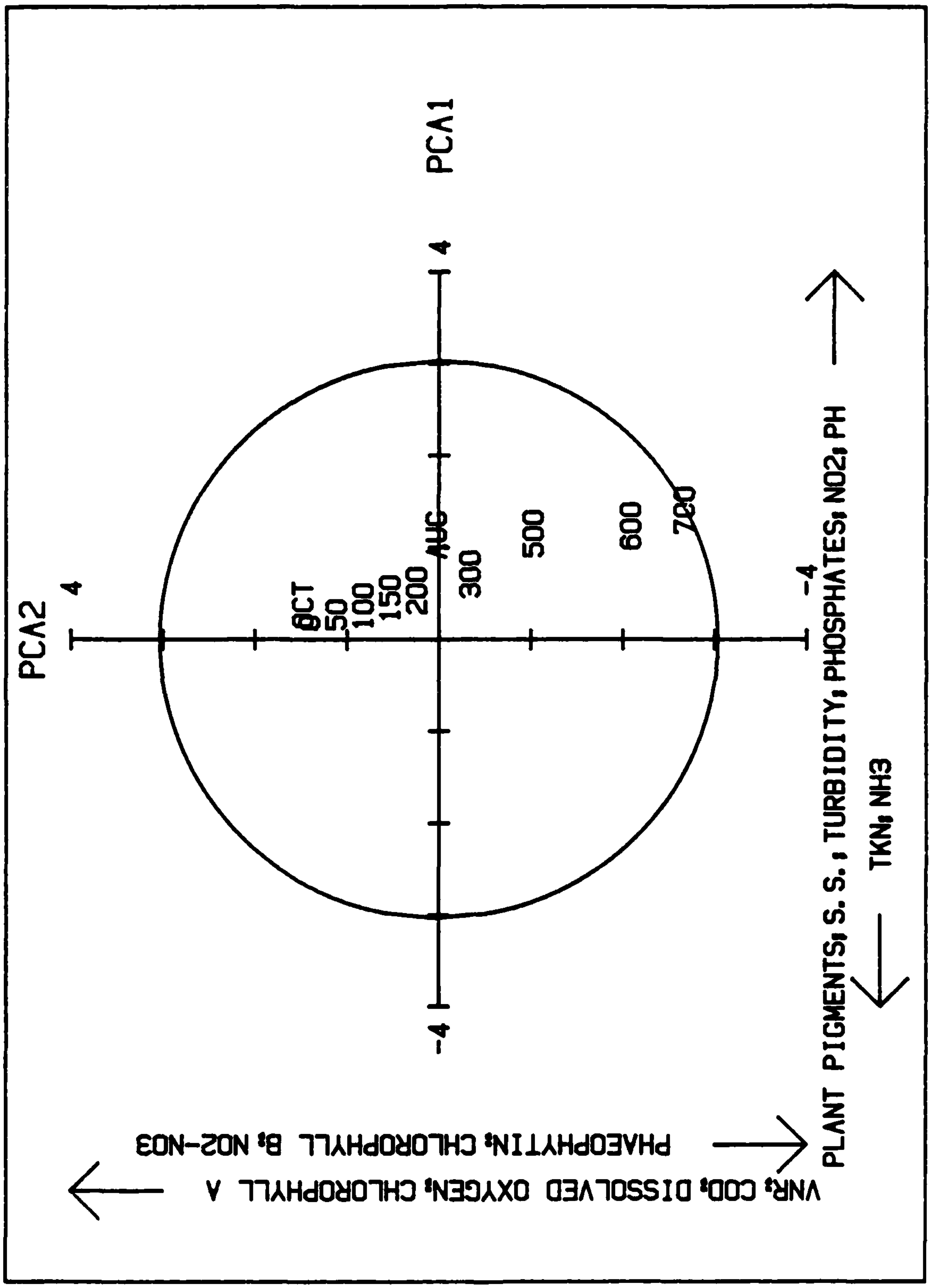
Empirical Tests of MDI Levels with Multivariate Models

Multivariate statistical techniques have been employed to characterize spatiotemporal water quality patterns of an ocean disposal site under baseline conditions (Alden et al., 1982). Similar methods can be employed in environmental monitoring studies to confirm the MDI levels during the baseline phase and, ultimately, to detect significant environmental impacts during the trend assessment phase. The techniques employed include principal components analysis (PCA), discriminant analysis, and multivariate analysis of variance (MANOVA). For demonstration purposes, the seasonal MDI levels presented in Table II will be evaluated using each of these techniques.

Green (1979) presents a method for assessing environmental impacts based upon a PCA of baseline data. Axes representing the first two factors, which account for most of the variance in the data, are plotted on a graph and a probability ellipse is calculated. The probability ellipse for any desired alpha level defines the boundary between a region of data points considered to be statistically similar to the baseline data and those considered to represent a significant impact. The PCA scores for each new sample or set of samples are calculated and evaluated with respect to the ellipse.

The means of various data sets plotted in relation to a probability ellipse defined by the set of baseline data from the Norfolk Disposal Site monitoring program are presented in Figure 3. The sets evaluated include data from the August and October cruises, as well as the October disposal site data which has been adjusted by factors of 50%, 100%, 200%, 300%, 500%, 600% and 700% of the estimated MDI values for each variable. The mean values of the simulated data sets do not exceed the 99% probability region until changed by factors of over seven times the MDI levels. The method appears to be quite insensitive to changes in the data. However, this is not too surprising since PCA tends to emphasize the patterns of variability

Figure 3. Mean values of standardized Principal Component Analysis (PCA) scores for August and October data and simulated data sets containing various levels of impact (50%, 100%, 200%, 300%, 500%, 600% and 700% of the MDI values for each variable) plotted on the 99% probability ellipse for the first two factors of a Principal Component Analysis (PCA) of the baseline data set. The relative loadings of the original variables on the axes are indicated.

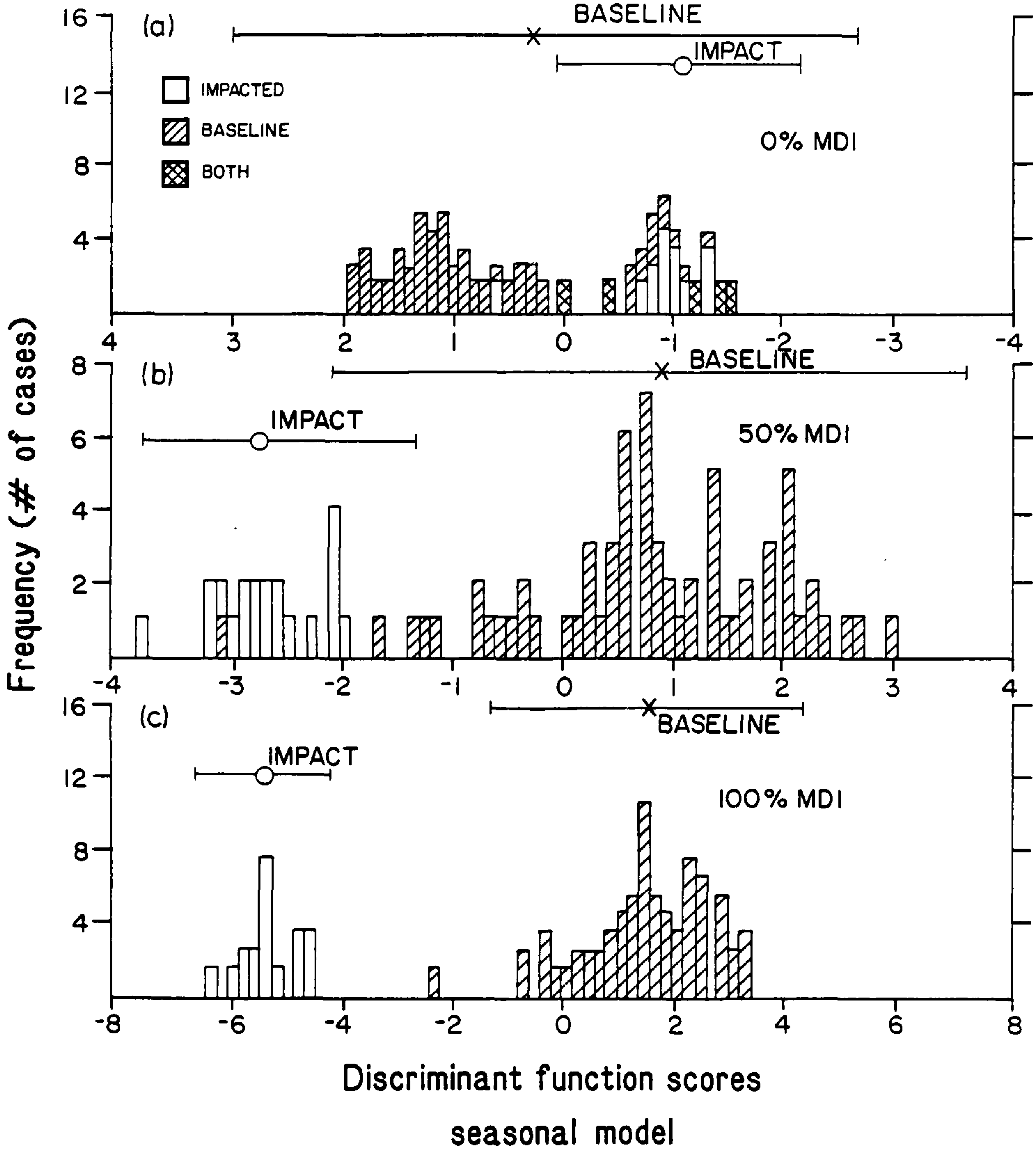

PCA1
PCA2
4
-4
4
-4
OCT
50
100
150
200
AUG
300
500
600
700
PLANT PIGMENTS; S. S. ; TURBIDITY; PHOSPHATES; NO2; PH
TKN; NH3
VNR; COD; DISSOLVED OXYGEN; CHLOROPHYLL A
PHAEOPHYTIN; CHLOROPHYLL B; NO2-NO3

within the data, so any impact must be relatively large before it becomes statistically significant under this model.

The PCA factors may represent combinations of variables which may be differentially affected by any given impact scenario. For example, the MDI's in Table II were calculated based upon the assumption that oxygen, pH and plant pigments would decrease following an impact, while the other variables would increase. Therefore, PCA factors loading on both "enhanced" and "inhibited" variables would not change very readily. This is the case for both factors (Figure 3). In addition, the relative "sensitivity" of the model (i.e. the ease with which a change is detected) depends upon the relative location of the season within the baseline ellipse. In other words, if an August cruise had been impacted, it would have been detected for a far smaller relative change from the ambient conditions. Although this phenomenon makes it more difficult to assess the MDI levels (which are on a relative scale) with this model, deviations in absolute concentrations of the variables from the baseline conditions would likely be detected.

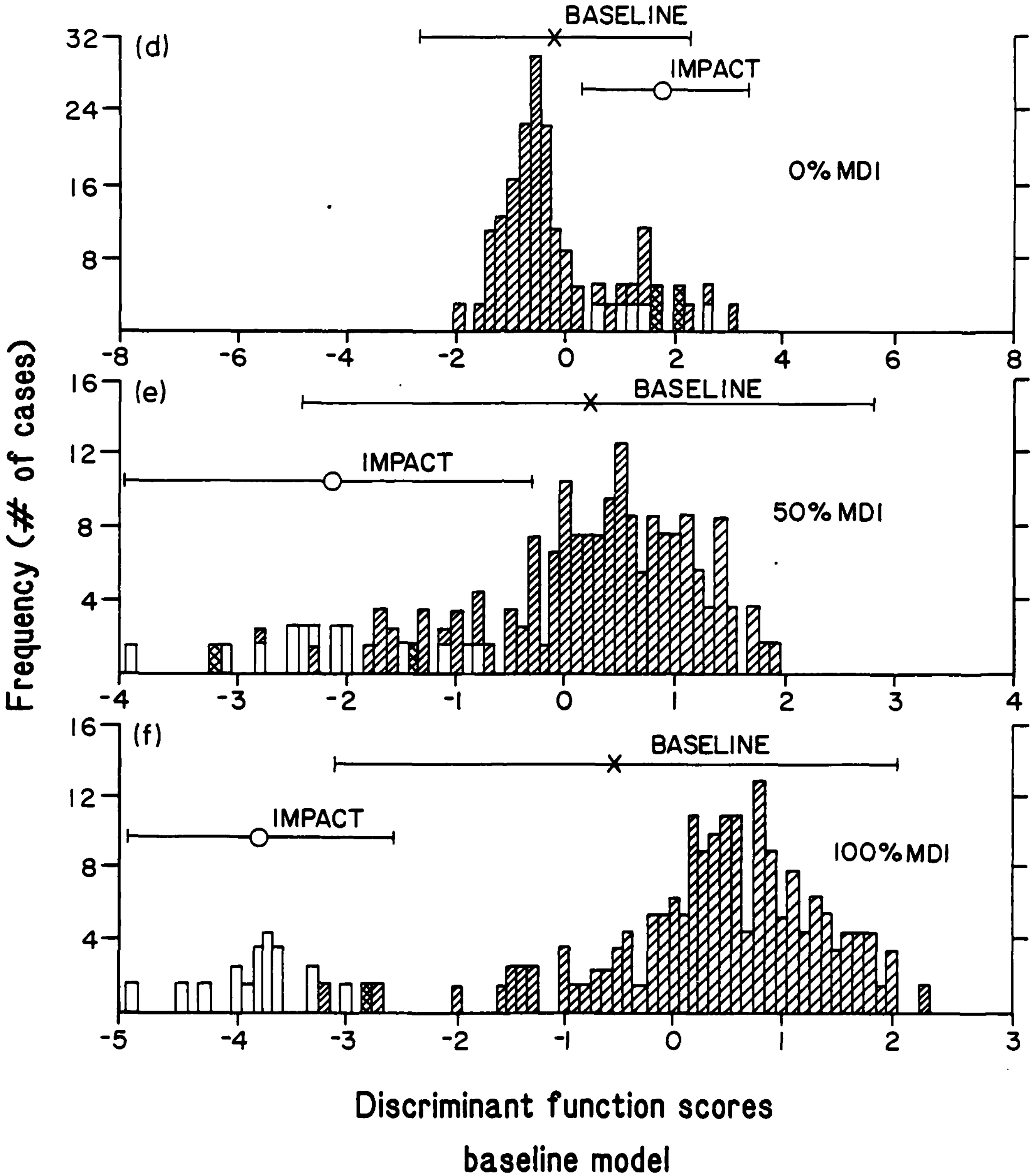
A second method involves the use of discriminant analysis in the comparison of "reference" with "impacted" data in either a baseline or seasonal model. A stepwise discriminant analysis procedure such as that described by Klecka (1975), forms a discriminant function based upon the variables which best separate the two groups. Figure 4 and Table III present the results of discriminant analysis of baseline and seasonal models with "impacted" data adjusted by factors which were 0%, 50%, and 100% of the MDI levels for each variable. Unlike the PCA models, the discriminant analyses are very sensitive. Unfortunately, the discriminant techniques are actually too sensitive: significant "impacts" were indicated even when none were introduced

Figure 4. Frequency histograms of discriminant function scores for baseline and "impacted" data: a) 0%, MDI seasonal model; b) 50% MDI seasonal model; c) 100%, MDI seasonal model.



(a)
BASELINE
IMPACT
IMPACTED
BASELINE
BOTH
0% MDI
(b)
BASELINE
IMPACT
50% MDI
(c)
BASELINE
IMPACT
100% MDI
Frequency (# of cases)
Discriminant function scores
seasonal model

Figure 4. Frequency histograms of discriminant function scores for baseline and "impacted" data: d) 0%, baseline model; e) 50% MDI baseline model and f) 100% MDI baseline model.



BASELINE
IMPACT
(d)
0% MDI
(e)
50% MDI
(f)
100% MDI
Frequency (# of cases)
-8 -6 -4 -2 0 2 4 6 8
-4 -3 -2 -1 0 1 2 3 4
-5 -4 -3 -2 -1 0 1 2 3
32 24 16 8
16 12 8 4
Discriminant function scores
baseline model

Table III. Summary of statistics from multivariate tests of data with simulated impacts versus data from baseline studies (see text for details of models and specific tests).

Model	Simulated Impact (% of Estimated MDI's)	Test	Criterion	Value	d.f.	P
I. Baseline	0%	Discriminant Analysis	Wilk's λ	0.731	1,6	< 0.0001
	50%	"	"	0.633	1,5	< 0.0001
	100%	"	"	0.344	1,11	< 0.0001
	0%	Discriminant Classification	% Correct Classification	60%		N/A
	50%	"	"	75%		"
	100%	"	"	100%		"
	0%	Manova	Multivariate F	4.72	14,173	< 0.0001
	50%	"	"	12.58	14,173	< 0.0001
	100%	"	"	23.64	14,173	< 0.0001
	0%	P(x/g) Classification of D.F. Scores	% Impacted scores outside of 99% control limits	5%		N/A
	50%	"	"	40%		"
	100%	"	"	100%		"
II. Seasonal	0%	Discriminant Analysis	Wilk's λ	0.818	1,6	< 0.0001
	50%	"	"	0.331	1,6	< 0.0001
	100%	"	"	0.111	1,5	< 0.001
	0%	Discriminant Analysis	% Correct Classification	75%		N/A
	50%	"	"	100%		"
	100%	"	"	100%		"
	0%	Manova	Multivariate F	1.36	15,76	0.200
	50%	"	"	10.85	15,76	< 0.0001
	100%	"	"	43.47	15,76	< 0.0001
	0%	P(x/g) Classification of D. F. Scores	% Impacted scores outside of 99% control limits	0%		N/A
	50%	"	"	75%		"
	100%	"	"	100%		"
III. Seasonal-Area Interaction	0%	Manova	Multivariate F	1.36	14,55	0.35
	50%	"	"	12.63	14,55	< 0.0001
	100%	"	"	38.32	14.55	< 0.0001
	0%	Univariate Tests Interaction Model Summary	% Test Significant at 0.05 level	7%		N/A
	50%	" "	"	43%		"
	100%	" "	"	100%		"

(0% MDI) (Table III). This was also not surprising since discriminant techniques emphasize differences between groups, even if they represent natural patterns. MANOVA models can be used to parallel the discriminant methods. A pre-disposal versus post-disposal effect is evaluated with respect to patterns in all variables with MANOVA techniques such as those described by Hull and Nie (1981). The results of these analyses are also indicated in Table III. Unlike the discriminant analysis, the MANOVA procedure did not indicate significant differences (at the $\alpha = 0.01$ level) between the "reference" and the 0% MDI data set for the seasonal model, although the mean of the 0% MDI data set was significantly different from the grand mean in the baseline model. The other "impacted" data sets (50% and 100% MDI) were shown to be very highly significantly different from "reference" sets for both models.

There is some question as to whether it is appropriate to utilize tests which statistically compare the mean of a post-impact data set with that of the total baseline or composite seasonal set. As demonstrated in Table III, it is entirely conceivable that the mean discriminant function or canonical variate value of data from any given cruise could be different from the grand mean due to natural variations alone. Therefore, a second test of significance was sought for the baseline and seasonal models.

The discriminant analysis procedure described by Klecka (1975) include a classification probability statistic, $p(x/g)$, which is defined as the probability that any given case belongs to the group with which it has been identified. The statistic is based upon a chi-square (see Figure 1):

$$\overline{X}^2_{[1]} = (f-\overline{f}_j)\ (D_j)^{-1}\ (f-\overline{f}_j)' \qquad (5)$$

where f is the discriminant score of the case being tested, $\bar{f}_j$ is the mean discriminant score for the group in question and D_j is the variance-covariance matrix, which reduces to the variance of scores for group j when there is a single discriminant function. The same statistic, which is analogous to that used by Green (1979) (equation 1) to calculate the probability ellipse, can be used to define probability limits for the discriminant scores of each group. The scores of each of the cases from the impact data set can then be evaluated by equation 5 to determine whether they would fall within the 99% probability limits of the baseline or seasonal "reference" data sets. This procedure was used for the data sets evaluated by the discriminant models, (Table III). The selection of a specific cut-off criterion in terms of the percentage of cases outside the probability limits being necessary to define an "impact" is somewhat subjective. A situation for which over 1% of the cases fall outside of the probability ellipse may be suspect. Once the mean discriminant function (D.F.) score of the post-impact data falls outside of the probability region of the baseline conditions, the majority of the cases in the data set are likely to do so as well, so this criterion has been tentatively adopted in the evaluation of MDI levels. The horizontal bars above the histograms in Figure 4 show the relationships between the mean D.F. scores and probability limits for each data set. As indicated by Table III, few of the 0% MDI cases fall outside of the probability limits for the "reference" groups, while all do for the 100% MDI data set, with the 50% MDI tests exhibiting intermediate results (e.g. 40% and 75% for seasonal and baseline models, respectively).

The MANOVA season-area interaction model essentially asks the question:

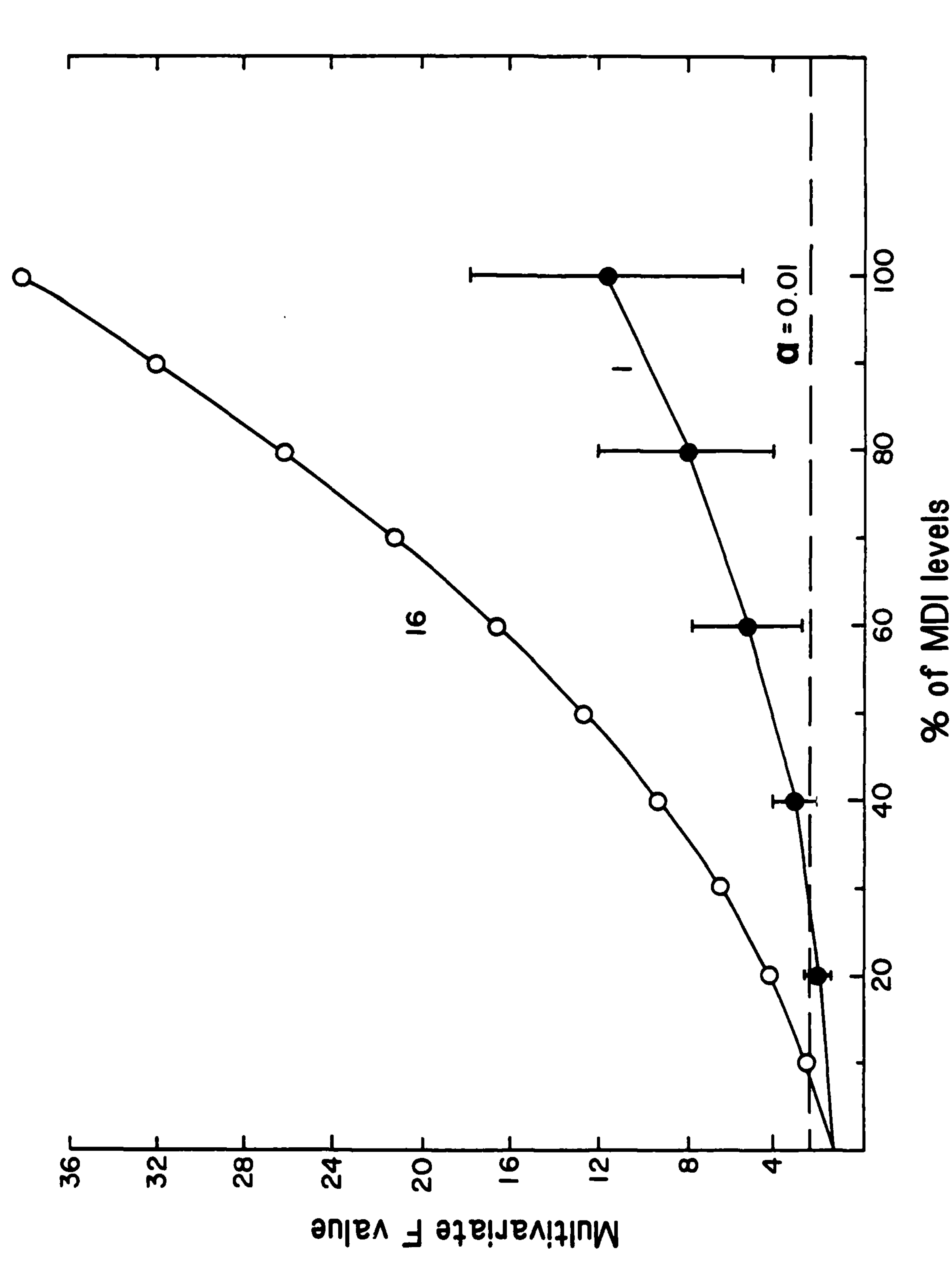
"Are there post-impact changes at the disposal site which are not observed elsewhere and are not accounted for by previously observed spatial differences?" The results of the MANOVA analysis of the August, October and simulate data sets are also presented in Table III. For the 0% MDI data set, the multivariate test showed no significant interaction effect ($p = 0.35$), while the univariate tests of each of the variables provided by the procedure indicated that no more were significant than would be expected by change alone. The 50% MDI data set was shown to produce a very highly significant multivariate interaction effect, while some of the variables were shown to be significant on their own. The interaction effect in the test of the 100% MDI set was significant for the multivariate test and all of the univariate tests of the individual variables. Thus, with the MANOVA interaction model, changes can be detected for impact levels below the predicted MDI's for the multivariate approach and at levels within the range of the expected MDI's for the univariate approach.

As with the case of the MDI's for single samples, it is conceivable, in fact quite likely, that all variables would not be impacted at the same time. Therefore, it would be useful to examine the relationship between the level of impact and the number of variables affected as they relate to the statistical significance of the test. A graph depicting an empirical relationship analogous to the one presented in Figure 1 for single samples would be desirable, but the large number of MANOVA tests involved would prove too time-consuming and costly in terms of computer usage to allow the evaluation of the same number of permutations. Therefore, tests involving various levels of impact for the extreme situations (i.e. one with a single impacted variable, or with all variables being impacted) were run under the assumption that other combinations would fall somewhere between the two curves. Figure 5 presents the results of this series of analyses, plotting the percent of the MDI level versus the multivariate F produced by the MANOVA procedure. The curve for the single impacted variable represents the means of ten tests of variables randomly selected from the data set to be adjusted by each given impact level. As with the results from the single sample tests, the curves appear exponential and it would appear that most situations involving combinations of impacted variables would require an average level of impact of 30-40% of the MDI's in order to be detected statistically by the MANOVA model. Although combinations of variables with different levels of impact may complicate the situation to a degree, it would seem intuitive that they should fall somewhere between the two empirical curves. Thus, it appears that, as for the single sample case, the predicted MDI levels represent estimators of statistically significant impacts which are conservative, but which are within the same order of magnitude as a change which would be detectable by a powerful multivariate statistical test.

EVALUATION OF STATISTICAL METHODS

One potential problem with the types of multivariate statistical models which have been described is the effect of heterogeneity of variances and non-normality on the test results. In order to evaluate this effect, a series of empirical tests were made utilizing data sets simulated to match the distribution of a highly skewed sediment grain size data set. The method of simulation is designed to produce data sets with desired distributions and various means (Alden, 1984). In the evaluation test series, discriminant analysis and MANOVA models were used to compare four groups of twenty cases which were simulated to have identical means (i.e. a true null

Figure 5. Multivariate F values versus the percent of the MDI's for a series of empirical tests. The curve for single impacted variables is based upon 10 tests of variables randomly selected from the data set to be adjusted for each impact level (means indicated by closed circles, vertical bars represent 2 standard errors). The curve labeled 16 is for all variables being impacted at once.


Multivariate F value
4
8
12
16
20
24
28
32
36
% of MDI levels
20
40
60
80
100
16
1
α = 0.01

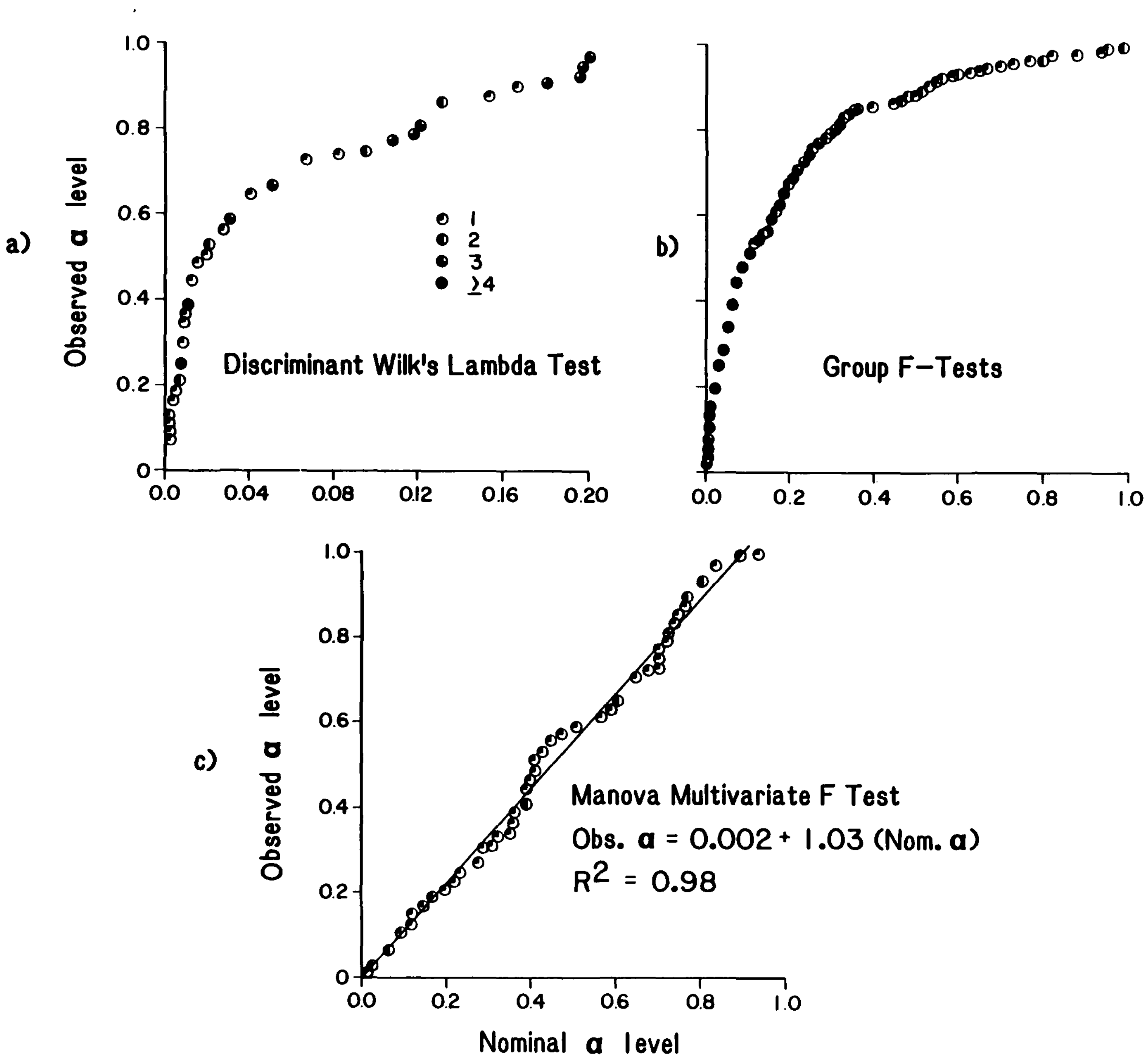
hypothesis). Tests were run for each type of analysis on fifty independent simulations. The nominal probability (α) values associated with the multivariate test statistics were then compared with the observed probability levels. The latter values were calculated as the cumulative proportion of the observations found for each nominal level.

The results of the comparisons of the observed versus nominal α levels for the multivariate tests are presented in Figure 6. An interesting contrast can be seen between the results of the discriminant analysis (Figure 6 a,b) and those from the MANOVA tests (Figure 6 c). The results from the discriminant analysis, Wilk's Lambda comparison of all groups and the Mahalanobis D^2 comparisons of individual pairs of groups (Klecka, 1975), clearly indicates a nonlinear relationship between the observed and nominal levels. Observed α values are always higher than nominal levels. Virtually all nominal α levels are below 0.20. In other words, the discriminant analysis models are overly sensitive, producing test statistics with "significant" nominal α levels, even when no true differences exist between the means.

On the other hand, the observed and nominal alpha levels for the MANOVA model were nearly equal throughout the entire range, producing a linear relationship of good fit ($R^2 = 0.99$). Therefore, the MANOVA models appear to produce expected results despite a great deviation from multivariate normality. However, if the linear relationship had not crossed the origin, the regression equation could have been used to predict the appropriate α level for any given test statistic (e.g. an "F" value producing a nominal α level of 0.05 may in fact be at an actual value of 0.07, etc.).

Between the results of the MDI evaluations and the simulation evaluations of the statistical, it is clear that MANOVA represents a "better" statistical model than does discriminant analysis in the comparison of en-

Figure 6. Observed versus nominal α levels for multivariate tests of data sets (50 runs of 4 groups each) simulated to fulfill a time null hypothesis: a) Discriminant Wilk's Lambda Test; b) Mahalanobis D^2 comparisons of individual pairs of groups; and c) MANOVA Multivariate F Tests.


a)
Observed α level
1.0
0.8
0.6
0.4
0.2
0
0.0
0.04
0.08
0.12
0.16
0.20
1
2
3
≥4
Discriminant Wilk's Lambda Test
b)
Group F—Tests
0.0
0.2
0.4
0.6
0.8
1.0
c)
Observed α level
Manova Multivariate F Test
Obs. α = 0.002 + 1.03 (Nom. α)
R2 = 0.98
Nominal α level

vironmental data sets. Discriminant analysis, while being overly sensitive as a test statistic, does have certain advantages: it can indicate which variables are most responsible for differences between groups; and the results can be readily displayed for presentation purposes (e.g. group distributions plotted on a frequency histogram of a single discriminant function, or probability ellipses of groups plotted on a graph of the first two discriminant functions). Therefore, results of a discriminant analysis may be used for data presentation purposes, but the "signficance" of the patterns should be confirmed with a MANOVA test.

CONCLUSIONS

A series of statistical approaches have been developed for the evaluation of baseline and trend assessment data. For baseline data sets, models have been developed which allow the estimation of "Minimum Detection Limits" for scenarios involving both single samples and entire data sets. The levels of MDI estimated for the data set selected for demonstration purposes indicated that the monitoring program was effective in providing an "early warning system" for the statistical detection of impacts before they become excessively detrimental to the environment. In fact, when more than one variable was affected at the same time, the amount of relative change in any given level was less (e.g. only 30-50% of MDI values for most multiple impacts). However, if the MDI's had represented conditions which would be considered to be damaging ecologically, models for the estimation of appropriate sample size for the detection of any desired level of change (e.g. see methods described by Sokal and Rohlf, 1969 and Green, 1979) could be used and computer simulations of data sets with this level of replication run with the MDI prediction methods to calculate new MDI's. Multivariate statistical models could be run to verify the new MDI levels.

A cost-benefit analysis could then be run to evaluate the relative amount of effort (and resources) required to potentially detect any given level of change (i.e. the "cost" of the various monitoring regimes could be plotted against the predicted MDI levels for evaluation purposes). In such an analysis, a point of "diminishing returns" could, perhaps, be observed and the optimum regime selected. At the very least, the investigator would be given an indication of the potential sensitivity of the regime which has been selected.

The effectiveness of several multivariate techniques for the detection of impacts in data from trend asssessment studies was evaluated empirically. A PCA model proved to be not very sensitive to the detection of significant changes, while discriminant analysis models were overly sensitive, suggesting significant differences when none were present. The MANOVA techniques proved to be the most effective, particularly when the season-area interaction models were used. Simulation evaluations confirmed the effectiveness and robustness of the MANOVA models. These simulation evaluations can also be used to define new critical test values for statistical models not exhibiting the nominal α values, thus freeing the statistical models from the assumptions of multivariate normality and homogeneity.

REFERENCES

Alden, R.W. III. 1984. A method for the simulation of multivariate environmental data. Accompanying report to the USACOE.

Alden, R.W. III, D.M. Dauer, and J.H. Rule. 1982. Environmental studies at a proposed mid-Atlantic dredged material disposal site. Oceans, 1034.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. Wiley-Interscience, N.Y. 257 pp.

Hull, C.H. and N.H. Nie. 1981. MANOVA: multivariate analysis of variance Chapter 1. In: C.H. Hull and N.H. Nie (eds.): Statistical Package for the Social Sciences, update 7-9, p. 1, McGraw-Hill, N.Y.

Kester, D.R. and R.A. Courant. 1973. A summary of chemical oceanographic conditions. In: Coastal and Offshore Environmental Inventory: Cape Hatteras to Nantucket Shoals. Univ. of R. I. Publication Series No. 2, Kingston, R.I.

Klecka, C. 1975. Discriminant analysis. Chapter 23, In: N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent (eds): Statistical Package for the Social Sciences, McGraw-Hill, N.Y. 434 pp.

Sokal, R.R. and F.J. Rohlf. 1969. Biometry: The principles and practices of statistics in biological research. W.H. Freeman and Co., San Franciso. 776 pp.

Virginia State Water Control Board. 1976. Water quality inventory (305(b) Report): Virginia Report to EPA Administration and Congress. Info. Bull. 526. 328 pp.

DTIC
FILMED
4-86
END